**MỘT PHƯƠNG PHÁP MỚI ĐỂ XÁC ĐỊNH GÁN XÁC SUẤT CƠ BẢN DỰA TRÊN ADABOOST VÀ ỨNG DỤNG CỦA NÓ TRONG PHÂN LOẠI**

**Các tác giả gốc: Wei Fu, Shuang Yu, Xin Wang, Éloi Bossé (biên tập viên học thuật)**

**Người dịch: Nguyễn Duy Thành (K64), Vương Tất Chiến (K64) - Trường Đại học Thủy Lợi**

Copyright © 2021 by Wei Fu, Shuang Yu, Xin Wang, Éloi Bossé. Published online in 2021 Jun 25 and available for everyone to read and research.

Vietnamese translation: Copyright © 2023 Duy Thanh, Vuong Tat Chien, Scientific research team and Thuy Loi University.

**This document for internal use only! Any copy or distribute without permission of Duy Thanh, Vuong Tat Chien, Scientific research team and Thuy Loi University are prohibited! Do not leak this document!**

**Last edited: 30/04/2023; 09:41 AM**

**I) Trừu tượng (Abstract):**

- Trong khuôn khổ của lý thuyết bằng chứng, một trong những vấn đề quan trọng và được mở ra là làm thế nào có thể xác định gán xác suất cơ bản (Basic probability assignment – BPA), điều mà trực tiếp liên quan tới kết quả quyết định có chính xác hay không. Bài báo này trình bày một phương pháp mới để có được gán xác suất cơ bản dựa trên Adaboost.

- Phương pháp này sử dụng dữ liệu huấn luyện để tạo ra các bộ phân loại mạnh mẽ một cách đa dạng cho mô hình thuộc tích, cái mà được sử dụng để xác định gán xác suất cơ bản của một đề xuất đơn lẻ kể từ khi sức mạnh của phân loại cung cấp thông tin hữu ích cho các giả thuyết cơ bản. Gán xác suất cơ bản (BPA) của một đề xuất tổng hợp được đo lường bởi việc tính toán tỉ lệ diện tích của một vùng giao điểm của đề xuất đơn lẻ. Công thức đệ quy của tỉ lệ diện tích của một vùng giao điểm đã được đề xuất và thật sự rất hữu ích cho tính toán bằng máy tính.

- Cuối cùng, phương pháp gán xác suất cơ bản được kết hợp cùng với quy tắc kết hợp của Dempster. Sử dụng phương pháp được đề xuất để phân loại tập dữ liệu Iris, các cuộc thử nghiệm đã có kết luận rằng tổng số lần nhận dạng thành công là 96.53% và khả năng phân loại trung bình đạt 90% khi mà quá trình huấn luyện đạt được 10%. Đối với các tập dữ liệu khác, các kết quả thử nghiệm cũng cho thấy phương pháp được đề xuất này là lí tưởng và có ảnh hưởng, và phương pháp được đề xuất này cũng thực hiện rất tốt trong trường hợp có các ví dụ không đủ.

**- Lưu ý: Kể từ đây trở về sau, phương pháp gán xác suất cơ bản sẽ được viết tắt là BPA(s), tức là Basic probability assignment (Phương pháp gán xác suất cơ bản).**

**I) GIỚI THIỆU**

- Trước đây, công nghệ tổng hợp thông tin đa cảm biến đã nhận được nhiều sự chú ý và được ứng dụng rộng rãi trong nhiều lĩnh vực, bao gồm trong quân đội, y tế và tài chính và còn nhiều lĩnh vực khác nữa. Trong khi nó tích hợp hiệu quả những thông tin từ nhiều nguồn khác nhau (các cảm biến), nó chủ yếu chọn những chiến thuật phù hợp. Một trong những điều quan trọng nhất được ủng hộ trong công nghệ kết hợp thông tin, lí thuyết bằng chứng DS (từ nay về sau sẽ được viết tắt là (DSET) có những lợi ích đáng kể. Nó không chỉ thể hiện các thông tin không chắc chắn một cách hiệu quả, mà còn có thể hợp nhất các thông tin mà không có các dữ liệu cho trước hoặc dữ liệu cho trước không đầy đủ. Vì vậy, nó được sử dụng rộng rãi trong nhiều ứng dụng trên nhiều lĩnh vực, bao gồm phân tích mối nguy hiểm, nhận dạng điểm, chẩn đoán lỗi và phân loại.

- Sự ảnh hưởng của quy tắc hợp nhất của Dempster phụ thuộc vào cách thức xây dựng có hợp lí hay không của phương pháp BPA (còn được gọi là chức năng khối lượng). BPA có cấu trúc dựa trên sự phân loại dữ liệu của nhiều cảm biến khác nhau và cung cấp sự hộ trợ cho một hoặc nhiều lớp mục tiêu. Nếu như BPA không phản ánh tốt những đặc trưng của mục tiêu, nó có thể tạo ra kết luận phản trực giác trong quá trình kết hợp BPA, đó không phải là kết quả được mong đợi. Vì vậy, để sử dụng DSET tốt hơn, tối thiểu phải tìm được con đường đi lý tưởng để xác định BPA. Nói chung, xác định BPA có thể được chia làm hai loại phương pháp:

+) Trong phương pháp đầu tiên, BPA được xác định bởi các chuyên gia. Mặc dù các ý kiến của các chuyên gia đưa ra chủ quan, việc xác định BPA đôi khi sẽ xảy ra tranh cãi mạnh.

+) Phương pháp thứ hai là điều khiển dữ liệu khi mà BPA tự động được xác định trong một số cách khác nhau

Bởi vì mức độ phức tạp và đa dạng của nền ứng dụng và sự gia tăng nhu cầu trong lý thuyết bằng chứng, xác định BPA trong DSET đề cập đến một số lượng lớn các vấn đề chưa được giải quyết và nó không có cách giải quyết tổng quát. Nhiều nhà nghiên cứu đã tạo ra vô số lần thử tại vấn đề này, nơi mà dựa trên hai lí thuyết: Lí thuyết xác định và lí thuyết mờ (fuzzy theory)

Dưới điều kiện của phân bố đồng đều, sự xác định BPA được dựa trên giả định rằng phân phối xác suất của các mẫu là mẫu Gaussian. BPA có nguồn gốc dựa trên sự phân phối phù hợp xác suất của các mẫu. Tiền đề của đa số các phương pháp được đề cập ở trên là các ví dụ huấn luyện có một sự phân phối cụ thể. Mặc dù, nền tảng của ứng dụng thực hành là rất phức tạp và đa dạng, và phân phối cụ thể của giả định chủ quan đôi khi không thể đáp ứng tình hình thực tế. Đặc biệt, trong các trường hợp mà các dữ liệu không cung cấp đầy đủ, càng khó để xác định phân phối của các dữ liệu huấn luyện. Vì vậy, một khi phân phối giả định bị sai lệch, rất khó để có được một kết quả mong muốn.

Phương pháp dựa trên lí thuyết mờ chủ yếu sử dụng số mờ tam giác để xác định BPA. Ví dụ, số mờ tam giác mở rộng được sử dụng để xác định BPA trong thế giới mở. Xiao đã đề xuất phương pháp tạo BPA, nơi mà thuật toán k-means được sử dụng để tối ưu hóa mẫu tạo BPA. Mặc dù vậy, cấu trúc của các hàm thành viên là khó khăn và là điểm mấu chốt trong việc sử dụng lý thuyết mờ để xác định BPA. Đặc biệt, hàm thành viên có thể sai lệch khi mà số lượng mẫu nhỏ, nó có thể làm giảm độ chính xác và tin cậy của BPA.

Như phương pháp điều khiển phân loại dữ liệu, Adaboost được sử dụng rộng rãi trong nhiều lĩnh vực, bao gồm phân loại, chẩn đoán lỗi, nhận diện điểm, dự đoán, năng lượng, hàng không và y tế. Adaboost không tạo ra bất cứ giả định nào về phân phối xác suất của các mẫu. Hơn nữa, nó đã một cấu trúc đơn giản và không dễ bị khớp quá mức (overfitting) dữ liệu huấn luyện. Vì vậy, các đặc điểm của thuật toán Adaboost rất phù hợp cho việc xác định BPA.

- Như được đề cập trong phần trên của cuộc thảo luận, bài báo này đề xuất phương pháp mới nhằm xác định BPA mà không có những giả định về phân phối xác suất cho mẫu, cái mà cũng sẽ bị ảnh hưởng khi số lượng mẫu nhỏ. Dựa trên các bước chính của thuật toán Adaboost, phương pháp xác định BPA bằng cách ghi lại đa số các phiếu bầu của các bộ phận loại yếu và sử dụng phương pháp tỉ lệ diện tích. Các bước như sau:

**+) Đầu tiên, một số bộ phận loại mạnh được sản sỉnh bằng cách áp dụng Adaboost.**

**+) Tiếp theo, BPA của đề xuất đơn lẻ được xác định bởi đa số phiếu bầu của các bộ phận loại yếu.**

**+) Bước thứ ba, phương pháp tỷ lệ diện tích được đề xuất được sử dụng để xác định BPA của đề xuất tổng hợp.**

**+) Cuối cùng, kết quả cuối cùng thu được bởi sự kết hợp tất cả các BPA sử dụng quy tắc kết hợp của Dempster.**

Những đóng góp chính của bài báo này như sau:

**(1): Phương pháp mới để xác định BPA dựa trên Adaboost được đề xuất, cái mà điều khiển dữ liệu và không tạo ra bất cứ giả định nào về phân phối xác suất cho mẫu, vì thế nó có thể làm giảm sự không chắc chắn của tính chủ quan**

**(2): Phương pháp tỉ lệ diện tích được đề xuất để xác định BPA của đề xuất tổng hợp, nó sẽ cải thiện khả năng đối phó với các thông tin không chắc chắn**

**(3): Phương pháp được đề xuất có tính phân loại tương đối cao và chính xác với số mẫu huấn luyện nhỏ**

**1.2) Cấu trúc tài liệu:**

- Cấu trúc của bài báo này như sau: **Phần 2** sẽ giới thiệu cơ bản về lí thuyết DSET và Adaboost. **Phần 3** sẽ giải thích về phương pháp được đề xuất và chi tiết kiến trúc của nó. **Phần 4**, thử nghiệm được thiết kế để đo lường sự ảnh hưởng của phương pháp được đề xuất. Cuối cùng, chúng tôi tổng kết lại kết quả và đưa ra kết quả trong **Phần 5**

**2) SƠ LƯỢC**

**2.1) SƠ LƯỢC VỀ LÍ THUYẾT DSET**

- So sánh với lí thuyết xác suất, **DSET** cung cấp công cụ mạnh mẽ để biểu hiện và kết hợp của các thông tin không chắc chắn mà không có xác suất trước, cái mà được sử dụng trong nhiều hệ thống hợp nhất dữ liệu. Một số điều cơ bản được trình bày dưới đây.

**+) Định nghĩa 1:**

**Khung phân biệt**

***“Nếu một tập hợp không rỗng Θ chứa tất cả kết quả của mục tiêu mà mọi người có thể nhận dạng, và mệnh đề được chứa trong tập hợp loại trừ lẫn nhau và toàn diện, nó được gọi là khung phân biệt”***



 ***là mệnh đề thứ i của khung nhận thức***

**+) Định nghĩa 2:**

***Gán xác suất cơ bản:***

***“Đặt  là khung riêng biệt, thì hàm***

******

***thỏa mãn các điều kiện sau:***

******

******

***Trong đó:  được gọi là BPA của A và được hiểu là thước đo lòng tin được cam kết chính xác với ***

**+) Định nghĩa 3:**

***Quy tắc kết hợp của Dempster:***

***“Giả sử rằng hai bằng chứng***  ***và của khung phân biệt  có các BPA  và  và  và  đại diện cho các yếu tố tiêu điểm khác nhau, tương ứng. Quy tắc kết hợp của Dempster sau đó được định nghĩa như sau:***





***Trong đó, k là hệ số xung đột và phản ánh mức độ xung đột giữa hai mảnh của bằng chứng. Nên lưu ý rằng thuật toán quy tắc kểt hợp của Dempster không thể được sử dụng với , có nghĩa là, khi đó hai mảnh sẽ xung đột với nhau hoàn toàn”***

**2.2) Adaboost**

- Adaboost là một thuật toán Boosting điển hình. Trong sự lặp đi lặp lại của huấn luyện, Adaboost tập trung nhiều hơn vào những mẫu chưa được phân loại và tạo ra những mẫu tương đối tốt khi kết thúc. Thuật toán không chỉ có cấu trúc đơn giản mà còn có độ chính xác cao.

- Adaboost không cần chọn các thuộc tính cho các mẫu huấn luyện. Thuật toán phân loại khác nhau có thể được sử dụng như những lớp loại yếu và có thể được xếp tầng trong các lớp loại mạnh. Trong thực tế, lớp loại yếu đơn giản nhất là gốc quyết định, nó là một dạng cây quyết định với một nhánh và thường được sử dụng trong các khuôn khổ của Boosting.

Trong khuôn khổ bài báo này, gốc quyết định được sử dụng như một lớp loại yếu. Có hai quá trình phân loại được đề cập dưới dạng một ví dụ để giới thiệu về quá trình huấn luyện của Adaboost:

**Bước 1:** Chọn tổ hợp huấn luyện:

****

Trong đó,  là số mẫu huấn luyện,

đại diện cho mẫu thứ i trong tổ hợp huấn luyện và  là lớp nhãn của 

**Bước 2:** Khởi tạo trọng số của tất cả mẫu huấn luyện:



Trong đó:

đại diện cho trọng số của  ở lần lặp đầu tiên

Bước 3: Huấn luyện lớp loại yếu  Đặt số lớn nhất của các lần lặp là  (số các lớp loại

yếu) và sau đó thực hiện quá trình huấn luyện của lần lặp  như sau:

(1): Huấn luyện lớp loại yếu  bằng cách sử dụng  và trọng số 



Trong đó  biểu diễn thuộc tính phần tử thứ d của  là giá trị ngưỡng cho thuộc tính cho thuộc tính thứ d trong vòng lặp, và  là hướng của một thuộc tính.

Nó có thể được nhìn thấy rằng quá trình phân loại của gốc quyết định là so sánh giá trị của  với  và nếu lớn hơn , thì  và . Ngược lại, β được sử dụng để sửa logic phán đoán khi  thu được kết quả phản trực giác

(2): Tính toán tỉ lệ lỗi  của các lớp loại yếu bằng:



Trong đó sẽ trả về hai giá trị 1 hoặc 0 và tương ứng với hai giá trị đó là các điều kiện cần được thỏa mãn như sau:



(3): Tính toán trọng số  của lớp loại yếu



(4): Cập nhật phân phối trọng số của mẫu:



Trong đó giá trị khởi đầu , x biểu thị cho bất cứ mẫu nào trong , và  là lớp nhãn của mẫu 

**Bước 4:** Lấy được lớp loại mạnh H bằng cách lặp lại **Bước 3:**

****

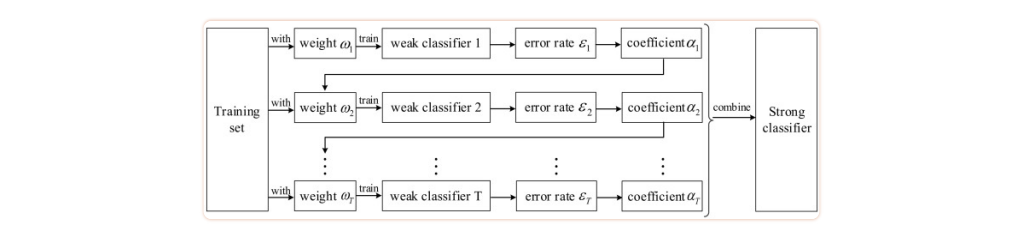
Trong đó,  là chức năng dấu hiệu:



Quá trình tính toán được ghi ở bảng sau:

|  |
| --- |
| **Thuật toán 1:** Quá trình tạo ra lớp loại mạnh dựa trên Adaboost |
| **Input:** Tập hợp huấn luyện D, số lượng lớp trong tệp dữ liệu:  , số lượng phân loại yếu:  **Output:**  : phân loại mạnh  **1: For** i = 1: – 1  **2: For** j = i + 1:  **3:** Thuộc tính i và j của tập huấn luyện ban đầu được chọn làm tập huấn luyện mới  **4:** Trọng số của mẫu huấn luyện được khởi tạo là .  **5: For** t = 1 : T  **6:** Sử dụng ωt của  , bộ phân loại yếu tối ưu được huấn luyện từ (6)  **7:** Tính toán tỉ lệ lỗi εt của  từ (7)  **8:** Dựa trên (9), tính toán trọng số  của  **9:** Dựa trên (10), sự phân bố trọng lượng  đã được cập nhật  **10: End For**  **11:** Dựa trên (11), nhận được phân loại mạnh  thuộc về thuộc tính của i và j  **12: End For**  **13: End For**  **14: Return** Tất cả các phân loại mạnh: |

Để hiểu hơn về thuật toán 1, hãy xem sơ đồ sau:



**3) Phương pháp được đề xuất để xác định BPA**

- Ở phần này, chúng tôi sẽ đề xuất phương pháp mới để xác định BPA dựa trên Adaboost. Đầu tiên, quá trình xác định BPA của đề xuất đơn lẻ được trình bày. Tiếp theo, phương pháp tỷ lệ diện tích được đề xuất để có được BPA của đề xuất tổng hợp

**3.1) Xác định BPA của đề xuất đơn lẻ**

- Dựa trên Adaboost được giới thiệu ở phần 2, tất cả các mẫu của hai thuộc tính trong tập huấn luyện được lấy như một tập huấn luyện mới. Người ta cho rằng các mẫu chứa lớp (), vì vậy chúng ta có thể có được phân loại mạnh , nơi mà phân loại thứ t

() được bao gồm bởi phân loại yếu 

Để có thể hiểu rõ hơn quá trình của phương pháp này, chúng ta giả định rằng A và B biểu thị cho lớp đầu tiên và lớp thứ hai của tập dữ liệu tương ứng. Nếu phân loại mạnh thứ t bầu cho mẫu thử nghiệm x, thì kết quả phiếu bầu của mẫu x thuộc về lớp A là:



Trong đó  là biểu thị cho phân loại yếu thứ k () của phân loại mạnh thứ t và  là trọng số của . Bởi vì số lớp của mẫu x có thể là A hoặc B trong quá trình bình chọn hiện tại, thì kết quả phiếu bầu của mẫu x thuộc về lớp B là:



Khi mà toàn bộ kết quả phiếu bầu của tất cả phân loại mạnh trong hai thuộc tính hiện tại được ghi lại, BPA của mẫu thử x thuộc về lớp A là:



Kết quả phiếu bầu ở trên được xác định dựa trên hai thuộc tính bất kì được chọn trong các mẫu. Nếu có thuộc tính n () trong các mẫu, BPA  được xác định. Tương tự, BPA của các lớp khác trong các mẫu thử có thể được xác định bằng công thức **(13) – (15)**

Phương pháp trong phần này chỉ xác định BPA trong thuyết đơn lẻ, đó là sự thiếu cân nhắc về tính không chắc chắn của mệnh đề tổng hợp. Để cải thiện khả năng của phương pháp xử lý thông tin không chắc chắn, chúng tôi đề xuất phương pháp tỷ lệ diện tích trong **phần 3.2**

**3.2. Xác định BPA của mệnh đề tổng hợp**

- Bài báo này giới thiệu phương pháp tỷ lệ diện tích để phân bổ lại các BPA của các mẫu nằm trong khu vực giao nhau. Kết quả của phân loại cho những mẫu này luôn luôn sai bởi vì rất khó có thể phân biệt các lớp mà chúng thuộc về, và đó là một loại thông tin không chắc chắn mà cần được thể hiện. Vì vậy, phương pháp tỷ lệ diện tích được đề xuất trong phần này miêu tả vùng giao nhau của các mẫu phân phối bằng cách xây dựng một vùng chữ nhật. Bằng cách tính tỉ lệ diện tích giữa các vùng, khối lượng của mệnh đề đơn lẻ là được phân bổ lại cho mệnh đề tổng hợp.

Để thuận tiện, chúng tôi giới thiệu ba kí hiệu:

(1): Trong bộ , mọi phần tử thỏa mãn  có thể được biểu diễn bởi kí hiệu chung . Ví dụ, trong khuôn khổ phân biệt , nếu hai quá trình phân loại hiện tại phân loại hai lớp A và B, thì  biểu thị A hoặc B, biểu thị thông tin không chắc chắn của AB, và biểu thị thông tin không chắc chắn của ABC

(2): Khi ,  biểu thị diện tích của vùng chữ nhật thuộc về các mẫu của mệnh đề riêng lẻ. Khi l > 1, biểu thị diện tích vùng giao nhau hình chữ nhật thuộc về các mẫu của l khác nhau trên các mệnh đề riêng lẻ. Ví dụ, chỉ vào diện tích của các vùng giao nhau hình chữ nhật A, B và C.

(3) Vùng tỉ lệ diện tích của  và  được định nghĩa như sau:



Trong quá trình phân bổ lại PBA, khối lượng của mệnh đề tổng hợp được lấy bởi quá trình đệ quy. Khối lượng  của sự không chắc chắn  được lấy bởi sự phân bổ lại theo tỷ lệ của khối lượng , nơi mà . Nó được diễn giải bằng công thức sau:



với giá trị khởi đầu là:

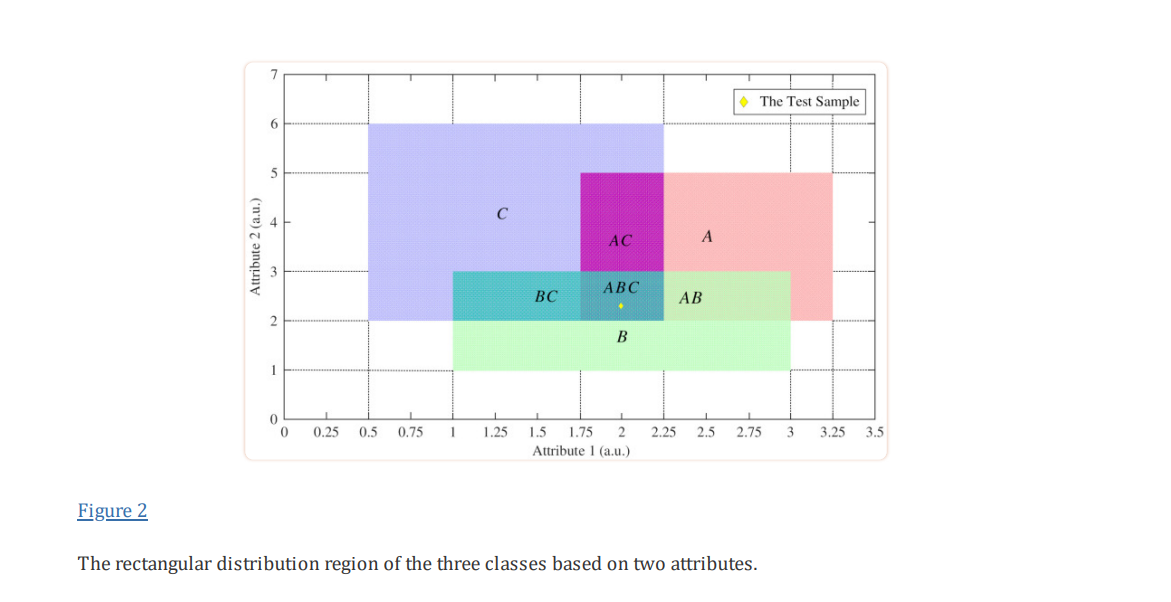


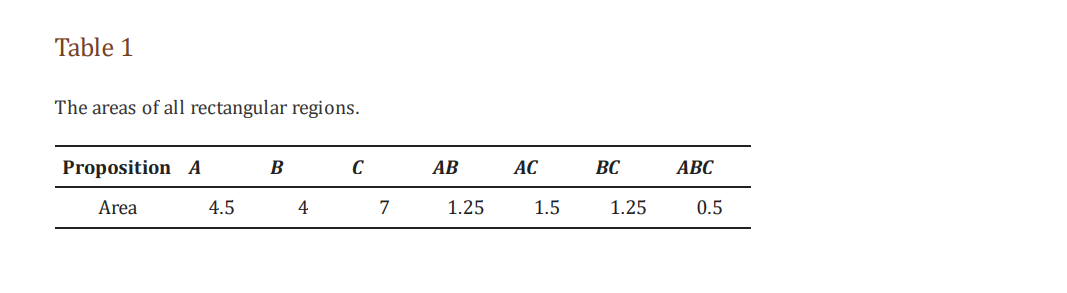
Kể từ khi khối lượng của mệnh đề đơn lẻ được phân bổ lại trở thành thông tin không chắc chắn theo tỉ lệ diện tích, khối lượng của sự tin tưởng không bị mất đi trong quá trình phân bổ lại. Các PBA được phân bổ lại thỏa mãn:



Ví dụ, đặt khung phân biệt bằng  và coi như các vùng mẫu của các lớp A, B và C giống như hình dưới đây. Diện tích của tất cả các vùng chữ nhật được lấy trong bảng dưới đây, nơi mà a.u. (1 a.u. = 149597871 km) là ngắn cho đơn vị tùy ý.

Hãy theo dõi biểu đồ và bảng sau đây:





Nếu khối lượng của mẫu được xác định bởi phân loại mạnh của hai lớp A và B, và mẫu thử được phân bổ trong vùng ABC, giống như hình dưới đây, quá trình phân bổ khối lượng của mệnh đề đơn lẻ theo phương pháp tỉ lệ diện tích như sau:



Quá trình diễn ra được miêu tả trong thuật toán 2:

|  |
| --- |
| **Thuật toán 2:** Phương pháp xác định PBA |
| **Input:**  Tập hợp huấn luyện: D, số lớp trong tập dữ liệu: , số lượng thuộc tính trong tập dữ liệu: n, số lần lặp: T  Tập hợp thử nghiệm P và số mẫu thử nghiệm .  **Output:**  Tất cả các BPA của tất cả các mẫu trong P  1: **For** i = 1: n – 1  2: **For** j = n + 1: n  3: Đặt một tổ hợp huấn luyện  mới dựa trên hai thuộc tính i và j của D  4: Đặt  vào Thuật toán 1 cho huấn luyện  phân loại mạnh  5: Đặt tổ hợp thử nghiệm  mới  6: **For** k = 1:  7: **For** t = 1:    8: nơi mà  là phân loại yếu của ,  là trọng số của  9:  10: **End For**  11:  12: **If**  thuộc bất cứ vùng giao nhau nào  13: Sử dụng công thức (17) để phân bổ lại khối  14: **End If**  15: **End For**  16: **End For**  17: **End For**  18: Hợp nhất tất cả BPA  sử dụng quy tắc kết hợp của Dempster để lấy được BPA , k = 1, …,  19: **Return** Tất cả BPA của tất cả các mẫu trong P |
|  |

**3.3: Cấu trúc của phương pháp được đề xuất**

- Trong phần này, quá trình xác định BPA dựa trên hai thuộc tính được mô tả chi tiết

**Bước 1:** Huấn luyện phân loại mạnh . Các mẫu của bất kì hai thuộc tính nào trong tổ hợp huấn luyện được lấy làm tổ hợp huấn luyện mới: , trong đó N là số lượng mẫu huấn luyện,  và  biểu thị dữ liệu thuộc tính thứ nhất và thứ hai của mẫu thứ i trong tổ hơp huấn luyện tương ứng, và  là nhãn lớp của . Thuật toán Adaboost trong phần 2.2 được sử dụng để huấn luyện phân loại mạnh  và khối lượng của tất cả phân loại yếu trong các phân loại mạnh được ghi lại

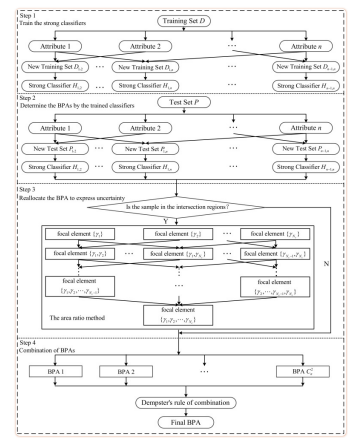
**Bước 2:** Xác định BPA bởi các phân loại được huấn luyện. Giống bước 1, các mẫu của bất kì hai thuộc tính nào trong tổ hợp mẫu huấn luyện được lấy làm tổ hợp mẫu huấn luyện mới

 trong đó  là số lượng các mẫu của các mẫu thử  và  biểu thị dữ liệu thuộc tính thứ nhất và thứ hai của mẫu thứ i trong tổ hơp mẫu huấn luyện tương ứng và  là nhãn lớp của . Chúng tôi sử dụng phân loại đã được huấn luyện trong bước 1 để bỏ phiếu cho từng mẫu trong tổ hợp thử, sau đó các BPA trong các mẫu thử được xác định bằng cách sử dụng Công thức **(13) – (15)**

**Bước 3:** Phân bổ lại BPA để biểu thị các thông tin không chắc chắn. Nếu mẫu  ở trong vùng giao nhau, sau đó chúng tôi sử dụng công thức **(17)** để phân bổ lại BPA được xác định ở Bước 2

**Bước 4:** Kết hợp các BPA. Từ bước 1 đến bước 3, Các BPA  được xác định cho từng mẫu thử, và chúng tôi có thể sử dụng quy tắc kết hợp của Dempster để tìm được BPA cuối cùng.

Để dễ hiểu, hãy theo dõi sơ đồ dưới đây:



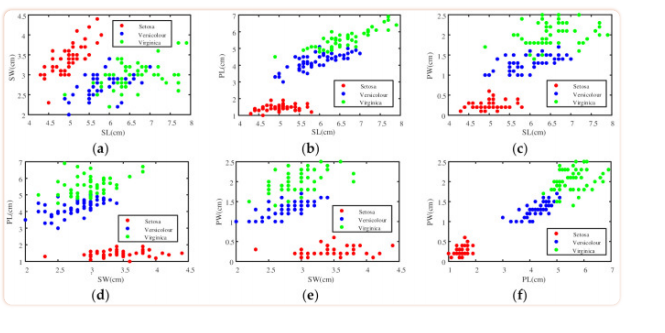
1. **Thí nghiệm**

Trong phần này, chúng tôi thiết kế một số thử nghiệm để chứng minh tính hiệu quả của phương pháp đề xuất về mặt phân loại và nhận dạng bằng cách sử dụng dữ liệu từ bộ dữ liệu học máy. Trong phần **4.1**, chúng tôi trình bày phương pháp được đề xuất với một ví dụ về xác định BPA bằng cách sử dụng bộ dữ liệu Iris. Trong phần **4.2,** chúng tôi sử dụng bốn bộ dữ liệu khác nhau để kiểm tra độ chính xác của phân loại và so sánh nó với độ chính xác phân loại của các phương pháp khác nhau.

* 1. **Ví dụ về bộ dữ liệu Iris để xác định BPA**

Bộ dữ liệu Iris được lấy từ kho lưu trữ học máy của UC Irvine, đây là một trong những bộ dữ liệu thường được sử dụng trong học máy. Bộ dữ liệu

Iris chứa ba lớp: Setosa (Se), Versicolour (Ve) và Virginica (Vi). Mỗi lớp chứa 50 mẫu và có bốn thuộc tính: chiều dài đài hoa (SL), chiều rộng đài hoa (SW), chiều dài cánh hoa (PL) và chiều rộng cánh hoa (PW). Theo phương pháp đề xuất trong bài báo này, bốn thuộc tính có thể được sử dụng để xác định sáu BPA của một mẫu thử nghiệm. Phân phối mẫu dựa trên hai thuộc tính được thể hiện trong **Hình 4**.



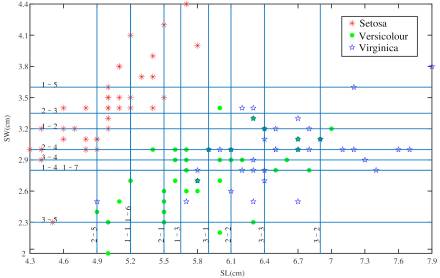
**Hình 4.**

Sáu số liệu phân phối dựa trên hai thuộc tính bất kỳ trong bộ dữ liệu Iris. ( **a** ) Phân phối mẫu dựa trên các thuộc tính SL và SW; ( **b** ) Phân phối mẫu dựa trên các thuộc tính SL và PL; ( **c** ) Phân phối mẫu dựa trên thuộc tính SL và PW; ( **d** ) Phân phối mẫu dựa trên các thuộc tính SW và PL; ( **e** ) Phân phối mẫu dựa trên các thuộc tính SW và PW; ( **f** ) Phân phối mẫu dựa trên các thuộc tính PL và PW.

**4.1.1. Xác định BPA của Singleton Proposition**

Trong thử nghiệm này, 40 nhóm mẫu được chọn ngẫu nhiên từ mỗi lớp của tập dữ liệu Iris, tổng cộng 120 mẫu được chọn làm tập dữ liệu huấn luyện và 30 mẫu còn lại được sử dụng làm tập dữ liệu kiểm tra. Theo dữ liệu của hai thuộc tính trong tập dữ liệu huấn luyện, một bộ phân loại mạnh được tạo ra, được sử dụng để bỏ phiếu cho các mẫu thử nghiệm nhằm xác định BPA. Các chi tiết của thí nghiệm được hiển thị dưới đây.

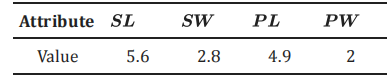
Một mẫu từ bộ thử nghiệm của Virginica được lấy làm ví dụ và dữ liệu được đưa ra trong **Bảng 2**. Vì tập dữ liệu huấn luyện chứa bốn thuộc tính nên chúng ta có thể nhận được sáu bộ phân loại mạnh. Dựa trên các mẫu đào tạo của SL và SW, **Hình 5** cho thấy các quy trình phân loại của phương pháp được đề xuất. Mỗi dòng trong biểu đồ đại diện cho một bộ phân loại yếu và ý nghĩa của số phía trên dòng có thể được mô tả là i−j, trong đó i đại diện cho quy trình phân loại hai lần thứ i và j đại diện cho trình phân loại yếu thứ j được đào tạo bởi quy trình phân loại hai lần thứ i.



**Hình 5.** Quá trình sử dụng bộ phân loại yếu để phân loại mẫu dựa trên SL và SW.

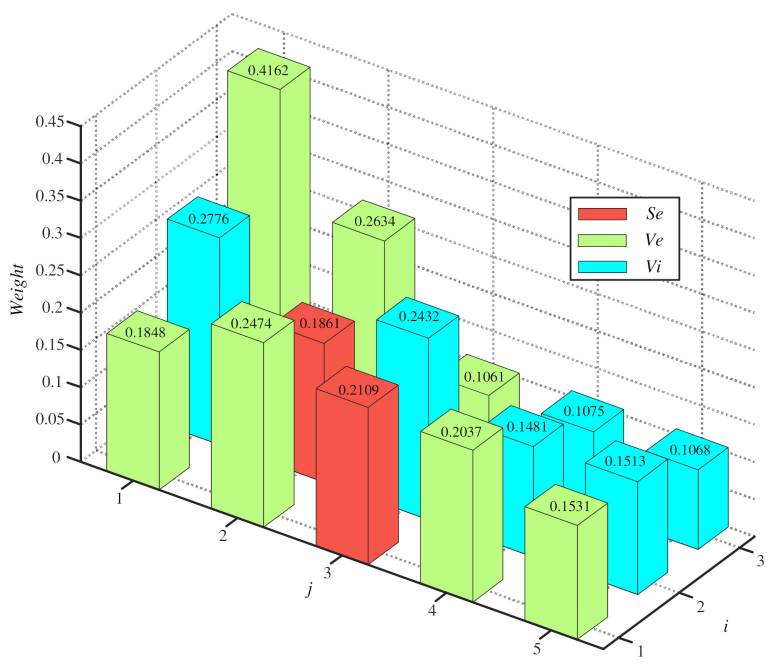
**Bảng 2.**

 Giá trị thuộc tính của mẫu.



Trọng số của các bộ phân loại yếu trong hai cách phân loại khác nhau được thể hiện trong **Hình 6**. Màu sắc khác nhau đại diện cho các lớp khác nhau và chiều cao đại diện cho giá trị của phiếu bầu. Bằng cách sử dụng Công thức **(13)–(15)** và phiếu bầu của tất cả các bộ phân loại yếu, khối lượng của mẫu này trong SL và SW được đưa ra như sau:

m(Se) = 0.1323, m(V e) = 0.5249, m(V i) = 0.3427

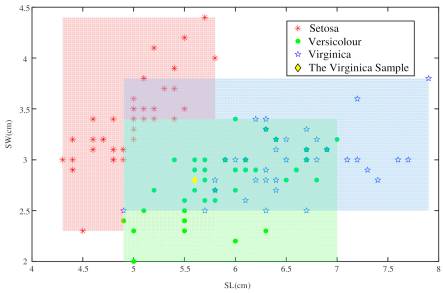


**Hình 6.** Trọng số của tất cả các bộ phân loại yếu dựa trên **Hình 5**.

Theo cách tương tự, chúng ta cũng có thể thu được kết quả biểu quyết của mẫu này ở hai thuộc tính bất kỳ khác. Tuy nhiên, mẫu Virginica đã cho nằm trong các vùng giao nhau của một số phân phối mẫu và nên xem xét tính không chắc chắn của mệnh đề tổng hợp. Do đó, chúng tôi đã sử dụng phương pháp tỷ lệ diện tích trong ví dụ này.

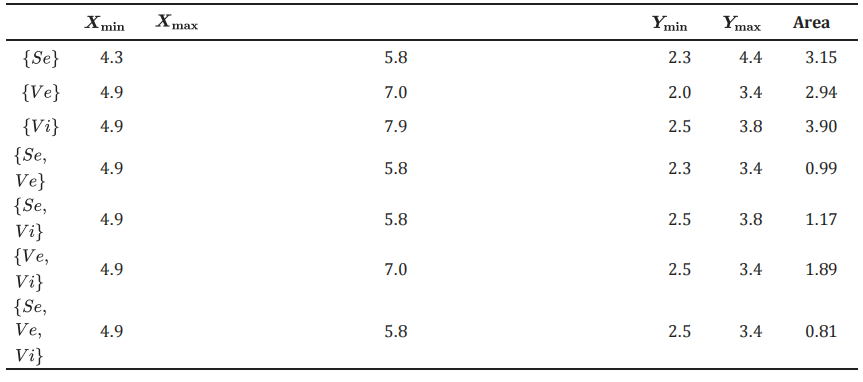
**4.1.2. Xác định BPA của Composite Proposition**

Như trong **Hình 7**, chúng tôi đã sử dụng các vùng hình chữ nhật với các màu khác nhau để biểu thị phạm vi phân phối của các phân phối mẫu cho các loại SL và SW khác nhau. Mẫu Virginica đã cho nằm trong khu vực giao nhau của ba khu vực phân phối. Do đó, chúng tôi đã tính toán tỷ lệ diện tích của các vùng giao nhau để chuẩn bị cho việc phân bổ lại BPA. Phạm vi và diện tích của tất cả các vùng trong thí nghiệm này được đưa ra trong **Bảng 3**.



**Hình 7.** Các vùng phân phối mẫu của tập dữ liệu huấn luyện dựa trên SL và SW.

**Bảng 3.** Phạm vi và diện tích của tất cả các vùng (cm).

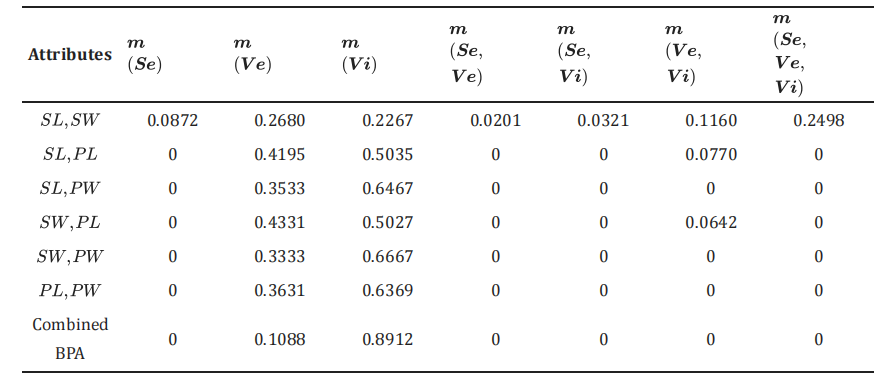


Bằng cách sử dụng các phương trình (17) và (18), chúng tôi nhận được:



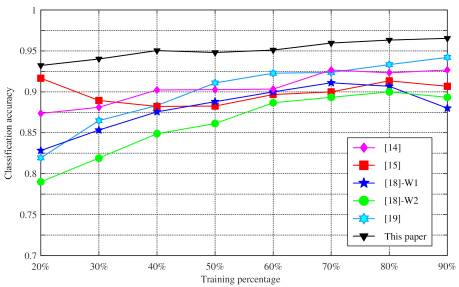
Từ phương trình (23) và (24), chúng ta có thể phân bổ lại kết quả biểu quyết của mẫu này với hai thuộc tính bất kỳ. Tất cả các kết quả biểu quyết được phân bổ lại và kết quả hợp nhất theo quy tắc kết hợp của Dempster được đưa ra trong **Bảng 4**.

**Bảng 4.** Tất cả các BPA được xác định và BPA kết hợp.



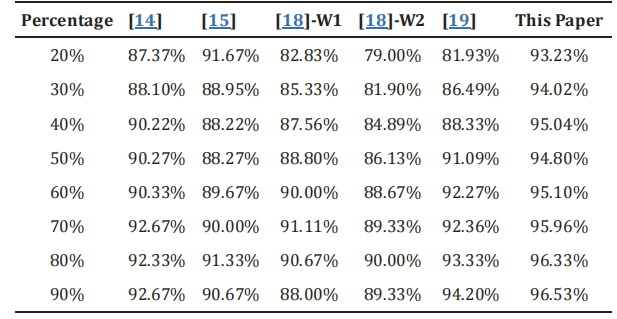
Từ các giá trị của BPA kết hợp trong **Bảng 4** , chúng ta có thể kết luận rằng loại của mẫu thử nghiệm là Virginica, phù hợp với kết quả của bộ dữ liệu Iris.

Để chứng minh tính ưu việt của phương pháp đề xuất, chúng tôi vẫn lấy tập dữ liệu Iris làm ví dụ để so sánh phương pháp đề xuất với phương pháp số khoảng [ **15** ] và phương pháp số mờ tam giác tổng quát [ **14** , **18** , **19** ]. Trong thí nghiệm này, số mẫu huấn luyện được chọn ngẫu nhiên từ mỗi lớp là 10, 15, 20, 25, 30, 35, 40 và 45. Các mẫu còn lại được dùng làm tập kiểm tra. Thí nghiệm được lặp lại 100 lần bằng phương pháp Monte Carlo và ghi lại giá trị trung bình của các kết quả thí nghiệm. Như trong **Hình 8** và **Bảng 5** , phương pháp được đề xuất trong bài báo này có độ chính xác phân loại cao hơn.



**Hình 8.** So sánh độ chính xác của bốn phương pháp khác nhau với phương pháp trong bài báo này.

**Bảng 5.** Độ chính xác so với phonon trăm đào tạo đối với các phương pháp khác nhau.



**4.2. Các thử nghiệm về việc thay đổi tỷ lệ phonon trăm đào tạo của bốn bộ dữ liệu UCI**

Trong phonon này, chúng tôi so sánh phương pháp được đề xuất với sáu bộ phân loại nổi tiếng sau: máy vector hỗ trợ (SVM), SVM với hàm cơ sở xuyên tâm (RBF), mạng RBF (RBFN), perceptron đa lớp (MP), naive Bayesian (NB) ) và Trình học cây quyết định (REPTree). Chúng tôi cũng xem xét Adaboost được đề cập trong **Phần 2.2** để minh họa tính hiệu quả của phương pháp được đề xuất. Ngoài bộ dữ liệu Iris, các thử nghiệm trong phonon này đã sử dụng ba bộ dữ liệu khác: Rượu, Viêm gan và Sonar, cũng từ kho lưu trữ máy học của UC Irvine.

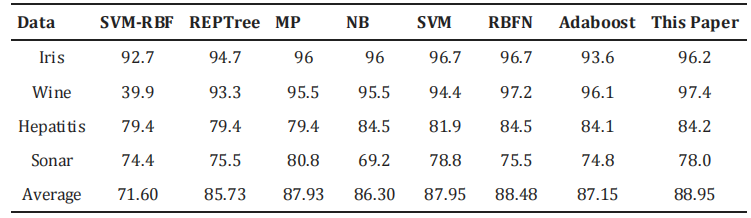
Bộ dữ liệu Rượu bao gồm 13 loại dữ liệu, là kết quả phân tích hóa học của ba loại rượu vang khác nhau được sản xuất trong cùng một vùng của Ý. Bộ dữ liệu Viêm gan chứa 19 thuộc tính, bao gồm thông tin bệnh nhân và kết quả xét nghiệm chức năng gan và dữ liệu của các thuộc tính này được sử dụng để dự đoán liệu bệnh nhân còn sống hay không. Bộ dữ liệu Sonar được sử dụng để dự đoán xem đối tượng mục tiêu là đá hay mỏ dựa trên dữ liệu cường độ do một sonar nhất định trả về từ các góc khác nhau. Thông tin cơ bản về các bộ dữ liệu này được đưa ra trong **Bảng 6**, bao gồm số lượng phiên bản, số lượng lớp, số lượng thuộc tính và tình trạng thiếu giá trị.

**Bảng 6.** Thông tin của các bộ dữ liệu khác nhau.

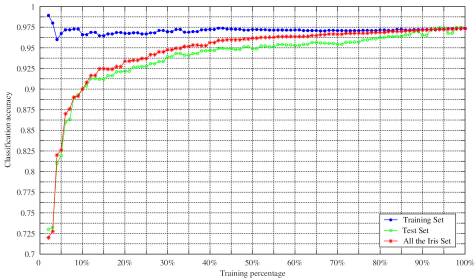


**Bảng 7** cho thấy dữ liệu về độ chính xác phân loại của các phương pháp phân loại khác nhau sử dụng bốn bộ dữ liệu trên. Trong thực nghiệm của mỗi phương pháp, 80% mẫu được chọn ngẫu nhiên làm tập dữ liệu huấn luyện và các mẫu còn lại làm tập dữ liệu kiểm tra. Sau đó, chúng tôi lặp lại thí nghiệm 100 lần và sử dụng độ chính xác trung bình của các thí nghiệm này làm độ chính xác cuối cùng. Bằng cách so sánh độ chính xác trung bình của từng phương pháp, có thể thấy phương pháp được đề xuất trong bài báo này hiệu quả hơn.

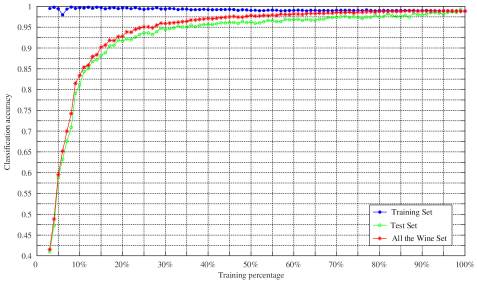
**Bảng 7.** Độ chính xác của các cách phân loại khác nhau (%).



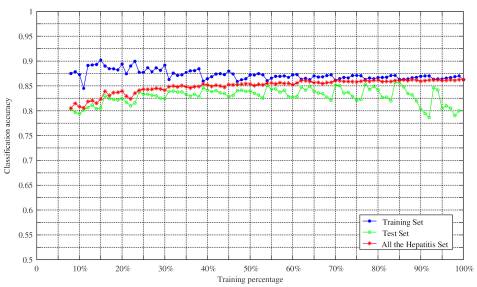
Để xác minh tính hiệu quả của phương pháp đề xuất trong phân loại, phương pháp đề xuất đã được thử nghiệm thêm bằng cách thay đổi tỷ lệ phần trăm đào tạo. N phần trăm mẫu của tập dữ liệu được chọn ngẫu nhiên làm tập huấn luyện và các mẫu còn lại được sử dụng làm tập kiểm tra. Chúng tôi đặt tỷ lệ phần trăm đào tạo của bộ dữ liệu Viêm gan từ 8% đến 98% vì nó chứa các giá trị bị thiếu, trong khi tỷ lệ phần trăm đào tạo của các bộ dữ liệu khác thay đổi từ 2% thành 98% trong quá trình đào tạo. Sau đó, phương pháp Monte Carlo được sử dụng để lặp lại thí nghiệm 100 lần để thu được độ chính xác phân loại trung bình của tập huấn luyện, độ chính xác phân loại trung bình của tập kiểm tra và độ chính xác phân loại trung bình của toàn bộ tập dữ liệu. Kết quả thí nghiệm được thể hiện trong **Hình 9** , **Hình 10**, **Hình 11** và **Hình 12** .



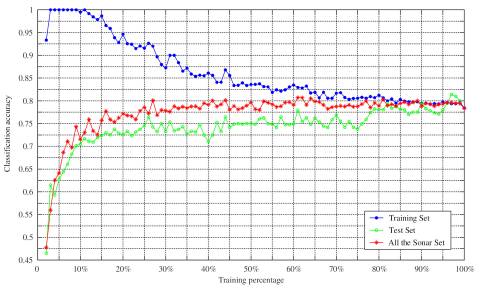
**Hình 9.** Độ chính xác phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Iris.



**Hình 10.** Độ chính xác của phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Rượu.



**Hình 11.** Độ chính xác của phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Viêm gan.

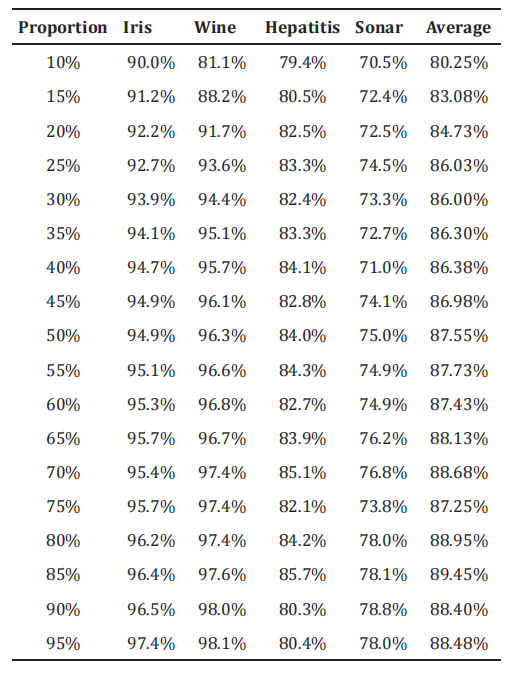


**Hình 12.** Độ chính xác phân loại so với tỷ lệ phần trăm đào tạo cho bộ dữ liệu Sonar.

Như có thể thấy từ **Hình 9** , **Hình 10**, **Hình 11** và **Hình 12** độ chính xác phân loại trung bình cho bộ dữ liệu Iris, bộ dữ liệu Rượu và bộ dữ liệu Sonar được cải thiện với tỷ lệ phần trăm đào tạo ngày càng tăng. Tuy nhiên, đối với bộ dữ liệu Viêm gan, xu hướng của độ chính xác phân loại trung bình không giống với các bộ dữ liệu khác và giảm khi số lượng mẫu thử nghiệm tăng lên. Điều này là do có 60 thuộc tính trong bộ dữ liệu Viêm gan và diện tích của vùng giao nhau giữa các thuộc tính khác nhau là lớn. Điều này gây khó khăn trong việc phân loại, đó là lý do tại sao hầu hết các thuật toán có độ chính xác phân loại tương tự nhau trong bộ dữ liệu Sonar. Tuy nhiên, độ chính xác phân loại trung bình của bộ dữ liệu Viêm gan vẫn tương đối cao.

Ngoài ra, trong lĩnh vực ứng dụng thực tế, có thể không thu được một số lượng lớn các mẫu đào tạo. Vì vậy, trong trường hợp này, tính khả thi của phương pháp xác định BPA là đặc biệt quan trọng. Như có thể thấy trong **Bảng 8** , độ chính xác của tập dữ liệu Iris và tập dữ liệu Sonar lần lượt đạt 81,28% và 90,26%, với tập huấn luyện là 10%. Khi tỷ lệ đào tạo là 15%, độ chính xác của bộ dữ liệu Rượu và bộ dữ liệu Viêm gan lần lượt là 88,2% và 80,5%. Điều đáng chú ý là độ chính xác phân loại trung bình của bốn bộ dữ liệu là 80,25% khi tỷ lệ đào tạo là 10%. Kết quả này cho thấy phương pháp trong bài báo này vẫn hợp lý và hiệu quả trong trường hợp số lượng mẫu huấn luyện nhỏ.

**Bảng 8.** Độ chính xác so với phần trăm đào tạo cho các bộ dữ liệu khác nhau.



1. **Kết luận**

Trong thuyết Dempster-Shafer (DSET), làm thế nào để xác định một phép gán xác suất cơ bản hợp lý (BPA), một bước quan trọng và đầu tiên, vẫn còn là một vấn đề mở. Trong bài báo này, một phương pháp mới để xác định BPA dựa trên Adaboost được đề xuất. Trong phương pháp được đề xuất này, nhiều bộ phân loại mạnh đã được xây dựng bằng cách sử dụng các mẫu huấn luyện và các trọng số tương ứng đã được ghi lại, được sử dụng để xác định BPA của mệnh đề đơn lẻ. BPA của mệnh đề tổng hợp được xác định bởi tỷ lệ diện tích vùng giao nhau của mệnh đề đơn lẻ. Ưu điểm của phương pháp đề xuất như sau:

1. Phương pháp được đề xuất trong bài báo này là dựa trên dữ liệu để giảm độ không chắc chắn do tính chủ quan gây ra.
2. Không có giả định nào được đưa ra về phân phối dữ liệu huấn luyện, điều này cho thấy có thể áp dụng phương pháp này trong nhiều lĩnh vực khác nhau.
3. Phương pháp tỷ lệ diện tích được đưa ra để cải thiện khả năng xử lý thông tin không chắc chắn của BPA ,đồng thời tăng độ chính xác của phân loại.
4. Đây là một phương pháp đơn giản, thiết thực và nó có thể xác định BPA trong trường hợp có số lượng mẫu đào tạo nhỏ.Sử dụng phương pháp được đề xuất để phân loại bộ dữ liệu Iris, thử nghiệm kết luận rằng tổng tỷ lệ nhận dạng là 96,53% và độ chính xác phân loại trung bình là 90% có thể đạt được khi tỷ lệ đào tạo là 10%.

Khi có quá nhiều thuộc tính của mẫu huấn luyện sẽ gây ra gánh nặng tính toán lớn hơn, đây chính là hạn chế của bài báo này. Là một phần mở rộng của các kết quả của bài báo này, các phương pháp xác định BPA dựa trên phân loại đa thuộc tính sẽ được xem xét trong công việc tương lai của chúng tôi.

## Sự nhìn nhận

Các tác giả đánh giá rất cao những gợi ý của các bài đánh giá và sự khuyến khích của biên tập viên.

## Sự đóng góp của tác giả

Khái niệm hóa, XW; phương pháp, WF; xác nhận, SY; phân tích chính thức, WF; quản lý dữ liệu, WF; viết—chuẩn bị bản thảo gốc, WF; viết—đánh giá và chỉnh sửa, WF, SY và XW; giám sát, SY; quản lý dự án, XW; mua lại tài trợ, XW Tất cả các tác giả đã đọc và đồng ý với phiên bản đã xuất bản của bản thảo.

## Kinh phí

Công trình này được hỗ trợ bởi Quỹ khoa học tự nhiên quốc gia Trung Quốc theo Khoản tài trợ số 61573132, Quỹ nghiên cứu cơ bản tại các tổ chức giáo dục đại học của tỉnh Hắc Long Giang theo Khoản tài trợ số KJCX201809 và RCYJTD201806.

## Xung đột lợi ích

Các tác giả tuyên bố không có xung đột lợi ích.

**Ghi chú của Nhà xuất bản:** MDPI giữ thái độ trung lập đối với các khiếu nại về quyền tài phán trong các bản đồ đã xuất bản và các cơ quan liên kết.

## Người giới thiệu và tài liệu tham khảo thêm

1. Chmielewski M., Kukiełka M., Pieczonka P., Gutowski T.Các phương pháp và công cụ phân tích để đánh giá tình hình chiến thuật trong các hoạt động quân sự bằng cách sử dụng cách tiếp cận tiềm năng và hợp nhất dữ liệu cảm biến. *Procedia Manuf.***2020** , *44* , 559–566.

doi: 10.1016/j.promfg.2020.02.255. [CrossRef] [Google Scholar]

1. Nagarani N., Venkatakrishnan P., Balaji N. Hệ thống hạ cánh trên đường băng của máy bay không người lái với khả năng phát hiện mục tiêu hiệu quả bằng cách sử dụng phản ứng tổng hợp hình thái cho hệ thống giám sát quân sự. *Comput. Commun.***2020** , *151* , 463–472.

doi: 10.1016/j.comcom.2019.12.039. [CrossRef] [Google Scholar]

3. Muzammal M., Talat R., Sodhro A.H., Pirbhulal S.Một cách tiếp cận tổng hợp dữ liệu đa cảm biến cho phép dữ liệu y tế từ các mạng cảm biến cơ thể.  *Inf. Fusion***2020** , *53* , 155–164.doi: 10.1016/j.inffus.2019.06.021. [CrossRef] [Google Scholar]

4.Magsi H., Sodhro A.H., Al-Rakhami M.S., Zahid N., Pirbhulal S., Wang L.Một thuật toán nhận biết pin thích ứng mới để truyền dữ liệu trong các ứng dụng chăm sóc sức khỏe dựa trên IoT.*Electronics* **2021** , *10* , 367.doi: 10.3390/electronics10040367. [CrossRef] [Google Scholar]

1. Jiang M.Q., Liu J.P., Zhang L., Liu C.Y.Một khung xếp chồng được cải thiện để dự đoán chỉ số chứng khoán bằng cách tận dụng các mô hình tập hợp dựa trên cây và thuật toán học sâu. *vật lý. Phys. A Stat. Mech. Its Appl.***2020** , *541* , 122272.

doi: 10.1016/j.physa.2019.122272. [CrossRef] [Google Scholar]

1. Himeur Y., Alsalemi A., Al-Kababji A., Bensaali F., Amira A.Các chiến lược hợp nhất dữ liệu để đạt hiệu quả năng lượng trong các tòa nhà: Tổng quan, thách thức và định hướng mới. *Inf. Fusion***2020** , *64* , 99–120.doi: 10.1016/j.inffus.2020.07.003. [CrossRef] [Google Scholar]
2. Makkawi K., Ait-Tmazirte N., el Najjar M.E., Moubayed N. Chẩn đoán thích ứng cho hợp nhất dữ liệu chịu lỗi dựa trên chiến lược phân kỳ α-Rényi để nội địa hóa phương tiện. *Entropy***2021** , *23* , 463.doi: 10.3390/e23040463. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
3. Li G.F., Deng Y.Một phép đo phân kỳ mới cho việc gán xác suất cơ bản và các ứng dụng của nó trong các môi trường cực kỳ không chắc chắn. *Int. J. Intell. Syst.***2019** , *34* , 584–600.

[Google Scholar]

1. Denœux T., Shenoy P.P.Một lý thuyết tiện ích có giá trị theo khoảng thời gian để ra quyết định với các hàm niềm tin Dempster-Shafer. *Int. J. Approx. Reason.***2020** , *124* , 194–216.

doi: 10.1016/j.ijar.2020.06.008. [CrossRef] [Google Scholar]

1. Song Y.T., Deng Y. Một phương pháp mới để đo lường sự khác biệt trong phản ứng tổng hợp dữ liệu cảm biến bằng chứng. *Int. J. Distrib. Sens. Netw.***2019** , *15* , 1–8.

doi: 10.1177/1550147719841295. [CrossRef] [Google Scholar]

1. Pan Y., Zhang L.M., Wu X.G., Skibniewski M.J. Hợp nhất thông tin đa phân loại trong phân tích rủi ro. *Inf. Fusion***2020** , *60* , 121–136.

doi: 10.1016/j.inffus.2020.02.003. [CrossRef] [Google Scholar]

1. Boukezzoula R., Coquin D., Nguyen T.L., Perrin S.Hợp nhất thông tin đa cảm biến: Sự kết hợp của các hệ thống mờ và phương pháp tiếp cận lý thuyết bằng chứng trong nhận dạng màu sắc cho rô-bốt hình người NAO. *Robot. Auton. Syst.***2018** , *100* , 302–316.

doi: 10.1016/j.robot.2017.12.002. [CrossRef] [Google Scholar]

13.Xiao Y.C., Xue J.Y., Zhang L., Wang Y.J., Li M.D. MD Chẩn đoán lỗi lệch trục cho tuabin gió dựa trên sự kết hợp thông tin. *Entropy***2021** , *23* , 243. doi: 10.3390/e23020243. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

14.Wu D.D., Liu Z.J., Tang Y.C. Một phương pháp phân loại mới dựa trên sự phủ định của phép gán xác suất cơ bản trong lý thuyết bằng chứng. *Eng. Appl. Artif. Intell.***2020** , *96* , 0952–1976.doi: 10.1016/j.engappai.2020.103985. [CrossRef] [Google

Scholar]

15.Kang B.Y., Li Y., Deng Y., Zhang Y.J., Deng X.Y.Xác định phép gán xác suất cơ bản dựa trên số khoảng và ứng dụng của nó. *Acta Electron. Sin.***2012** , *40* , 1092–1096.[Google Scholar]

1. Xu P.D., Deng Y., Su X.Y., Mahadevan S. Một phương pháp mới để xác định việc gán xác suất cơ bản từ dữ liệu huấn luyện. *Knowl. Based Syst***2013** , *46* , 69–80.

doi: 10.1016/j.knosys.2013.03.005. [CrossRef] [Google Scholar]

17.Chen H.F., Wang X.Xác định phép gán xác suất cơ bản dựa trên phân phối xác suất. Trong Kỷ yếu Hội nghị Kiểm soát Trung Quốc lần thứ 39 năm 2020, Thẩm Dương, Trung Quốc, ngày 9 tháng 9 năm 2020; trang 2941–2945.[Google Scholar]

18.Xiao J.Y., Tong M.M., Zhu C.J., Wang X.L. Phương pháp xây dựng gán xác suất cơ bản dựa trên số mờ tam giác tổng quát. *Chin. J. Sci. Instrum.***2012** , *32* , 191–196.[Google Scholar]

19.Zhang J.F., Deng Y.Một phương pháp để xác định việc gán xác suất cơ bản trong thế giới mở và ứng dụng của nó trong phân loại và hợp nhất dữ liệu. *Appl. Intell.***2017** , *46* , 934–951.

doi: 10.1007/s10489-016-0877-9. [CrossRef] [Google Scholar]

1. Jiang W., Zhan J., Zhou D., Li X.Một phương pháp để xác định phân bổ xác suất cơ bản tổng quát trong thế giới mở. *Math. Probl. Eng***2016** , *2016* , 3878634.

doi: 10.1155/2016/3878634. [CrossRef] [Google Scholar]

21.Fan Y., Ma T.S., Xiao F.Y. Một cách tiếp cận cải tiến để tạo ra phép gán xác suất cơ bản tổng quát dựa trên các tập mờ trong thế giới mở và ứng dụng của nó trong hợp nhất thông tin đa nguồn. *Appl. Intell.***2020** , *51* , 3718–3735.doi: 10.1007/s10489-020-01989-6. [CrossRef] [Google Scholar]

22.Li L., Wang C.Y., Li W., Chen J.B.Phân loại hình ảnh siêu phổ bằng các máy học cực trị hạt nhân tổng hợp có trọng số adaboost. *Neurocomputing***2018** , *275* , 1725–1733.

doi: 10.1016/j.neucom.2017.09.004. [CrossRef] [Google Scholar]

1. Liu H., Zhang X.C., Zhang X.T.huật toán adaboost dựa trên thế giới khả thi để phân loại dữ liệu không chắc chắn. *Knowl. Based Syst***2019** , *186* , 104930.

doi: 10.1016/j.knosys.2019.104930. [CrossRef] [Google Scholar]

24.Tang D., Tang L., Dai R., Chen J.W., Li X., Rodrigues J.J.P.C. MF-Adaboost: Phát hiện tấn công LDoS dựa trên đa tính năng và cải tiến adaboost.  *Future Gener. Comput. Syst.***2020** , *106* , 347–359.doi: 10.1016/j.future.2019.12.034. [CrossRef] [Google Scholar]

25.Li J.L., Sun L.J., Li R.N. Phát hiện không phá hủy thời gian chiên dầu đậu nành bằng công nghệ quang phổ NIR với adaboost-SVM (RBF). *Optik***2020** , *206* , 164248.

doi: 10.1016/j.ijleo.2020.164248. [CrossRef] [Google Scholar]

26.Wu Y.L., Ke Y.T., Chen Z., Liang S.Y., Zhao H.L., Hong H.Y. Ứng dụng cây quyết định xen kẽ với nhóm adaboost và đóng bao để lập bản đồ tính nhạy cảm trượt lở đất. *CATENA***2020** , *187* , 104396.doi: 10.1016/j.catena.2019.104396. [CrossRef] [Google Scholar]

27.Hu G.S., Yin C.J., Wan M.Z., Zhang Y., Fang Y. Công nhận các cây Pinus bị bệnh trong hình ảnh UAV bằng cách sử dụng trình phân loại học sâu và adaboost. *Biosyst. Eng.***2020** , *194* , 138–151.doi: 10.1016/j.biosystemseng.2020.03.021. [CrossRef] [Google

Scholar]

28.He Y.L., Zhao Y., Hu X., Yan X.N., Zhu Q.X., Xu Y. Chẩn đoán lỗi bằng cách sử dụng phép chiếu bảo tồn cục bộ phân biệt đối xử dựa trên adaboost mới với các mẫu lại. *Eng. Appl. Artif. Intell.***2020** , *91* , 103631.doi: 10.1016/j.engappai.2020.103631. [CrossRef] [Google Scholar]

29.. Zhao B., Zhang X.M., Li H., Yang Z.B.Chẩn đoán lỗi vòng bi thông minh dựa trên CNN đã chuẩn hóa xem xét sự mất cân bằng dữ liệu và điều kiện làm việc thay đổi.*Knowl. Based Syst.***2020** , *199* , 105971.doi: 10.1016/j.knosys.2020.105971. [CrossRef] [Google Scholar]

30.Messai O., Hachouf F., Seghir Z.A. Mạng thần kinh Adaboost và chế độ xem cyclopean để đánh giá chất lượng hình ảnh lập thể không tham chiếu. *Signal Process. Image Commun.***2020** , *82* , 115772.doi: 10.1016/j.image.2019.115772. [CrossRef] [Google Scholar]

31.Agbele T., Ojeme B., Jiang R. Ứng dụng các mẫu nhị phân cục bộ và trình phân loại adaboost theo tầng để phát hiện và phân tích các mẫu hành vi của chuột. *Procedia Comput. Sci.***2019** , *159* , 1375–1386.doi: 10.1016/j.procs.2019.09.308. [CrossRef] [Google Scholar]

32.Lin G.C., Zou X.J. Citrus cho máy thu hoạch tự động kết hợp với bộ phân loại adaboost và ngân hàng bộ lọc Leung-Malik. *IFAC-Pap.***2018** , *51* , 379–383.doi: 10.1016/j.ifacol.2018.08.192. [CrossRef] [Google Scholar]

33.Yang H., Liu S.L., Lu R.X., Zhu J.Y. Dự đoán hàm lượng thành phần trong quy trình khai thác đất hiếm dựa trên ESNs-adaboost. *IFAC-Pap.***2018** , *51* , 42–47.[Google Scholar]

34. Li H.H., Liu S.S., Hassan M.M., Ali S., Ouyang Q., Chen Q.S., Wu X.Y., Xu Z.L.Phân tích định lượng nhanh dư lượng Hg2+ trong các sản phẩm sữa bằng SERS kết hợp với thuật toán ACO-BP-adaboost. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.***2019** , *233* , 117281.doi: 10.1016/j.saa.2019.117281. [PubMed] [CrossRef] [Google Scholar]

35.Asim K.M., Idris A., Iqbal T., Martínez-Álvarez F.Hệ thống dự đoán động đất dựa trên các chỉ số địa chấn sử dụng lập trình di truyền và phân loại adaboost. *Soil Dyn. Earthq. Eng***2018** , *111* , 1–7.doi: 10.1016/j.soildyn.2018.04.020. [CrossRef] [Google Scholar]

36.Xiao C.J., Chen N.C., Hu C.L., Wang K., Gong J.Y., Chen Z.Q.Dự đoán nhiệt độ bề mặt nước biển ngắn hạn và trung hạn sử dụng dữ liệu vệ tinh chuỗi thời gian và phương pháp kết hợp LSTM-Adaboost. *Remote Sens. Environ.***2019** , *233* , 111358. doi: 10.1016/j.rse.2019.111358. [CrossRef] [Google Scholar]

37.Sun W., Gao Q. Khám phá tiềm năng tiết kiệm năng lượng trong ngành điện Trung Quốc dựa trên mạng thần kinh lan truyền ngược adaboost. *J. Sạch sẽ. sản xuất.***2019** , *217* , 257–266.

doi: 10.1016/j.jclepro.2019.01.205. [CrossRef] [Google Scholar]

38.Xu X.B., Duan H.B., Guo Y.J., Deng Y.M. Một thuật toán adaboost và CNN theo tầng để phát hiện giả mạo trong quá trình tiếp nhiên liệu trên không tự động của UAV. *Neurocomputing***2020** , *408* , 121–134.

doi: 10.1016/j.neucom.2019.10.115. [CrossRef] [Google Scholar]

39.Jiménez-García J., Gutiérrez-Tobal G.C., García M., Kheirandish-Gozal L., Martín-Montero A., Álvarez D., del Campo F., Gozal D., Hornero R. Đánh giá luồng không khí và tín hiệu đo oxy để phát hiện hội chứng ngưng thở khi ngủ ở trẻ em bằng cách sử dụng adaboost. *Entropy***2020** , *22* , 670.

doi: 10.3390/e22060670. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

**Ghi nhận bản quyền**

**Các tác giả gốc: Wei Fu, Shuang Yu, Xin Wang, Éloi Bossé (biên tập viên học thuật)**

**Người dịch: Nguyễn Duy Thành (K64), Vương Tất Chiến (K64) - Trường Đại học Thủy Lợi**

Copyright © 2021 by Wei Fu, Shuang Yu, Xin Wang, Éloi Bossé. Published online in 2021 Jun 25 and available for everyone to read and research.

Vietnamese translation: Copyright © 2023 Duy Thanh, Vuong Tat Chien, Scientific research team and Thuy Loi University.

**This document for internal use only! Any copy or distribute without permission of Duy Thanh, Vuong Tat Chien, Scientific research team and Thuy Loi University are prohibited! Do not leak this document!**

**Last edited: 30/04/2023; 09:41 AM**